

NUMERICAL METHODS IN THE WEAK SENSE FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH SMALL NOISE*

G. N. MILSTEIN[†] AND M. V. TRET'YAKOV[‡]

Abstract. We propose a new approach to constructing weak numerical methods for finding solutions to stochastic systems with small noise. For these methods we prove an error estimate in terms of products $h^i \varepsilon^j$ (h is a time increment, ε is a small parameter). We derive various efficient weak schemes for systems with small noise and study the Talay–Tubaro expansion of their global error. An efficient approach to reducing the Monte-Carlo error is presented. Some of the proposed methods are tested by calculating the Lyapunov exponent of a linear system with small noise.

Key words. weak approximation, small noise, computer simulation, Monte-Carlo methods

AMS subject classifications. 60H10, 65U05

PII. S0036142996278967

1. Introduction. Numerical methods adapted to specific systems can clearly be more efficient than general methods. An important instance of a stochastic system is given by a stochastic differential equation with small noise, since often fluctuations, which affect a dynamical system, are sufficiently small.

In [9] we studied the mean-square approximation of stochastic differential equations with small noise. Mean-square numerical methods are themselves significant. Moreover, they are the basis for the construction of weak schemes. In many cases numerical methods in the weak sense (cf. [5]–[8], [12, 14]) are appropriate for solving physical problems by a Monte-Carlo technique. They are easier to implement than mean-square methods and the required random variables can be efficiently simulated. However, for general systems, weak methods of more than second order tend to require laborious calculations. Besides, there are no sufficiently efficient high-order weak Runge–Kutta schemes.

Below we systematically study weak numerical methods for a stochastic system with small noise,

$$(1.1) \quad dX = a(t, X)dt + \varepsilon^2 b(t, X)dt + \varepsilon \sum_{r=1}^q \sigma_r(t, X) dW_r(t), \quad X(t_0) = X_0,$$

where $t \in [t_0, T]$, $0 \leq \varepsilon \leq \varepsilon_0$, ε is a small parameter, ε_0 is a positive number, $X = (X^1, X^2, \dots, X^n)$, $a(t, x) = (a^1(t, x), \dots, a^n(t, x))$, $b(t, x) = (b^1(t, x), \dots, b^n(t, x))$, $\sigma_r(t, x) = (\sigma_r^1(t, x), \dots, \sigma_r^n(t, x))$, $r = 1, \dots, q$, are n -dimensional vectors, $W_r(t)$, $r = 1, \dots, q$, are independent standard Wiener processes, and X_0 does not depend on $W_r(t) - W_r(t_0)$, $t_0 < t \leq T$, $r = 1, \dots, q$.

The errors of the proposed methods will be estimated in terms of products $h^i \varepsilon^j$, where h is the step-size of the discretization. Usually global error estimates are of the form $|E[f(X(T)) - f(\bar{X}(T))]| = O(h^p + \varepsilon^k h^q)$, $q < p$, where f is a function from a

*Received by the editors December 21, 1994; accepted for publication (in revised form) April 1, 1996.

<http://www.siam.org/journals/sinum/34-6/27896.html>

[†]Weierstraß-Institut für Angewandte Analysis und Stochastik, Mohrenstr., 39, D 10117 Berlin, Germany (milstein@wias-berlin.de).

[‡]Department of Mathematics, Ural State University, Lenin St., 51, 620083 Ekaterinburg, Russia (michael.tretyakov@usu.ru). The research of this author was supported in part by the Deutscher Akademischer Austauschdienst with a Forschungsstipendium aus Sondermitteln der NATO.

sufficiently large class and $\bar{X}(T)$ is an approximation of the exact solution $X(T)$. The time-step order of such a method is equal to q , which may be small. For small ε the product $\varepsilon^k h^q$ will also become small, and consequently, so will the global error. This allows us to construct effective weak methods with low time-step order but which nevertheless have small errors.

2. Main theorem on estimating a method's global error.

DEFINITION 2.1. A function $f(x)$ belongs to the class \mathbf{F} , $f \in \mathbf{F}$, if there are constants $K > 0$ and $\kappa \geq 0$ such that the inequality

$$(2.1) \quad |f(x)| \leq K(1 + |x|^\kappa)$$

holds for all $x \in \mathcal{R}^n$. A function $f(s, x)$, which depends both on $x \in \mathcal{R}^n$ and on a parameter $s \in Q$, belongs to the class \mathbf{F} (with respect to x) if the inequality (2.1) holds uniformly with respect to $s \in Q$.

Note that below, the same letter K is used for various constants, and similarly, $K(x)$ is used for various functions, $K(x) \in \mathbf{F}$.

THEOREM 2.1. Assume that the following conditions hold.

(1) The coefficients of the system (1.1) are continuous and satisfy a Lipschitz condition with respect to $x \in \mathcal{R}^n$; they and their partial derivatives up to a sufficiently high order belong to the class \mathbf{F} .

(2) The error of a one-step approximation $\bar{X}_{t,x}(t+h)$ of the exact solution $X_{t,x}(t+h)$ of the system (1.1) with initial condition $X(t) = \bar{X}(t) = x$ can be estimated by

$$(2.2) \quad |Ef(X_{t,x}(t+h)) - Ef(\bar{X}_{t,x}(t+h))| \leq K(x)[h^{p+1} + \sum_{l \in S} h^{l+1} \varepsilon^{J(l)}], K(x) \in \mathbf{F},$$

where the function $f(x)$ and its partial derivatives up to a sufficiently high order belong to the class \mathbf{F} , S is a subset of the positive integers $\{1, 2, \dots, p-1\}$, and J is a decreasing function from S to the positive integers.

(3) For a sufficiently large integer m the moments $E|\bar{X}_k|^m$ exist and are uniformly bounded with respect to N , $k = 0, 1, \dots, N$, and $0 \leq \varepsilon \leq \varepsilon_0$.

Then for any N and $k = 0, 1, \dots, N$,

$$(2.3) \quad |Ef(X_{t_o, X_o}(t_k)) - Ef(\bar{X}_{t_o, X_o}(t_k))| \leq K \left[h^p + \sum_{l \in S} h^l \varepsilon^{J(l)} \right],$$

where the constant K depends on the function $f(x)$, on the random variable X_o , and on T .

The proof of Theorem 2.1 differs only little from the proof of the weak convergence theorem given in [7, 8] and is therefore omitted.

Remark. If a method is such that it satisfies the inequality (2.2) with $p > 0$, then Theorem 2.1 implies the convergence of the method. However, the primary concern of Theorem 2.1 is to estimate the global error of a method in terms of h and ε .

3. Construction of weak methods for a system with small noise. Returning to the system (1.1) we introduce an equidistant discretization Δ_N of the interval $[t_o, T]$: $\Delta_N = \{t_k : k = 0, 1, \dots, N; t_o < t_1 < \dots < t_N = T\}$; the time increment $h = t_{k+1} - t_k$; the approximation X_k , \bar{X}_k , or $\bar{X}(t_k)$ of the exact solution $X(t_k)$; operators

$$L = L_1 + \varepsilon^2 L_2,$$

$$L_1 = \frac{\partial}{\partial t} + \left(a, \frac{\partial}{\partial x} \right) = \frac{\partial}{\partial t} + \sum_{i=1}^n a^i \frac{\partial}{\partial x^i},$$

$$L_2 = \left(b, \frac{\partial}{\partial x}\right) + \frac{1}{2} \sum_{r=1}^q \left(\sigma_r, \frac{\partial}{\partial x}\right)^2 = \sum_{i=1}^n b^i \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{r=1}^q \sum_{i,j=1}^n \sigma_r^i \sigma_r^j \frac{\partial^2}{\partial x^i \partial x^j},$$

$$\Lambda_r = \left(\sigma_r, \frac{\partial}{\partial x}\right) = \sum_{i=1}^n \sigma_r^i \frac{\partial}{\partial x^i};$$

and Ito integrals

$$I_{i_1, \dots, i_j}(t, h) = \int_t^{t+h} dW_{i_j}(\vartheta) \int_t^{\vartheta} dW_{i_{j-1}}(\vartheta_1) \int_t^{\vartheta_1} \dots \int_t^{\vartheta_{j-2}} dW_{i_1}(\vartheta_{j-1}),$$

where i_1, \dots, i_j are from the set of numbers $\{0, 1, \dots, q\}$ and $dW_o(\vartheta_i)$ stands for $d\vartheta_i$.

We assume that restrictions on the coefficients of the system (1.1) are imposed, which ensure the existence and uniqueness of the solution on the whole time interval $[t_o, T]$. For the construction of higher-order methods, the coefficients must be sufficiently smooth functions.

According to Theorem 2.1, to estimate the global error of a method we need properties of the corresponding one-step approximation; i.e., to prove an error for a weak method, we need the estimate (2.2). By using the Taylor expansion of the function f it is possible to obtain the estimate (2.2), provided the inequalities

$$(3.1) \quad \left| E \prod_{j=1}^m \Delta^{i_j} - E \prod_{j=1}^m \bar{\Delta}^{i_j} \right| \leq K(x) \left(h^{p+1} + \sum_{l \in S} h^{l+1} \varepsilon^{J(l)} \right), \quad m = 1, \dots, s-1,$$

$$\Delta^{i_j} = X^{i_j}(t+h) - x^{i_j}, \quad \bar{\Delta}^{i_j} = \bar{X}^{i_j}(t+h) - x^{i_j}, \quad X(t) = \bar{X}(t) = x, \quad i_j = 1, \dots, n,$$

$$(3.2) \quad E \prod_{j=1}^s |\bar{\Delta}^{i_j}| \leq K(x) \left(h^{p+1} + \sum_{l \in S} h^{l+1} \varepsilon^{J(l)} \right), \quad i_j = 1, \dots, n,$$

$$(3.3) \quad E \prod_{j=1}^s |\Delta^{i_j}| \leq K(x) \left(h^{p+1} + \sum_{l \in S} h^{l+1} \varepsilon^{J(l)} \right), \quad i_j = 1, \dots, n,$$

hold. The number s in (3.1)–(3.3) must be such that $h^k(\varepsilon h^{1/2})^{s-k} = O(h^{p+1} + \sum_{l \in S} h^{l+1} \varepsilon^{J(l)})$ for $k = 0, 1, \dots, s$.

To construct a one-step approximation, one frequently uses a truncated expansion of the exact solution in terms of Ito integrals (for the stochastic Taylor-type expansion, see [5], [8], [17]). On the basis of such an expansion, a one-step approximation $\tilde{X}(t+h)$ of weak order 3 can be derived for a general system ($\varepsilon = 1$), which in the case of the system (1.1) has the form

$$\begin{aligned} \tilde{X}(t+h) = & x + \varepsilon \sum_{r=1}^q \sigma_r I_r + h(a + \varepsilon^2 b) + \varepsilon^2 \sum_{i,r=1}^q \Lambda_i \sigma_r I_{ir} + \varepsilon \sum_{r=1}^q (L_1 + \varepsilon^2 L_2) \sigma_r I_{or} \\ & + \varepsilon \sum_{r=1}^q \Lambda_r (a + \varepsilon^2 b) I_{ro} + h^2 (L_1 + \varepsilon^2 L_2) (a + \varepsilon^2 b) / 2, \end{aligned}$$

$$(3.5) \quad X(t+h) = \tilde{X}(t+h) + \tilde{\rho}.$$

The coefficients σ_r , a , b , $\Lambda_i \sigma_r$, etc. in (3.4) are calculated at the point (t, x) , and $\tilde{\rho}$ in (3.5) is the remainder.

On the basis of the approximation (3.4) the following weak method with global error $O(h^2)$ was derived in [7, 8] (see also [5]):

$$\begin{aligned} X_{k+1} = & X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k + \varepsilon^2 h \sum_{i,r=1}^q (\Lambda_i \sigma_r \xi_{ir})_k \\ & + \varepsilon h^{3/2} \sum_{r=1}^q ((L_1 + \varepsilon^2 L_2) \sigma_r (\xi_r - \eta_r))_k + \varepsilon h^{3/2} \sum_{r=1}^q (\Lambda_r (a + \varepsilon^2 b) \eta_r)_k / 2 \\ & + h^2 ((L_1 + \varepsilon^2 L_2)(a + \varepsilon^2 b))_k / 2, \end{aligned} \quad (3.6)$$

where the independent random variables are simulated according to

$$\begin{aligned} \eta_r = \xi_r / 2, \quad \xi_{ir} = (\xi_i \xi_r - \gamma_{ir} \zeta_i \zeta_r) / 2, \quad \gamma_{ir} = \begin{cases} -1, & i < j, \\ 1, & i \geq j, \end{cases} \\ P(\xi = 0) = \frac{2}{3}, \quad P(\xi = -\sqrt{3}) = P(\xi = \sqrt{3}) = \frac{1}{6}, \\ P(\zeta = -1) = P(\zeta = 1) = \frac{1}{2}. \end{aligned} \quad (3.7)$$

A higher-order method would have to be based on a one-step approximation of weak order 4, which would be too complicated. However, as will be shown below, a small modification of the one-step approximation (3.4) leads to an efficient method with the local error $O(h^4 + \varepsilon^2 h^3)$.

The remainder $\tilde{\rho}$ in (3.5) contains terms with factors $\varepsilon^3 h^{3/2}$, $\varepsilon^2 h^2$, $\varepsilon^4 h^2$, $\varepsilon h^{5/2}$, $\varepsilon^3 h^{5/2}$, $\varepsilon^5 h^{5/2}$. Contribution of these terms to the local error of the one-step weak approximation is not worse than $O(\varepsilon^2 h^3)$, which follows from properties of Ito integrals (see [5, 7, 8]). Additionally, $\tilde{\rho}$ contains the term $h^3 L_1^2 a / 6$, which yields a contribution to the local error of the one-step weak approximation equal to $O(h^3)$, and terms, for example, $\varepsilon^2 h^3$, h^4 , etc., which contribute to the local error not more than $O(\varepsilon^2 h^3)$. It is clear that moving the term $h^3 L_1^2 a / 6$ from $\tilde{\rho}$ to \tilde{X} leads to a new one-step approximation which is not essentially more complicated, but is at the same time considerably more accurate than the previous one. Of course, this preliminary reasoning requires a strict proof. The proof is given below and is a consequence of the results in [7, 8].

We consider a new approximation \tilde{X} (here, since it can be done without causing ambiguity, we use the same notation \tilde{X} and $\tilde{\rho}$ for the new one-step approximation and its remainder):

$$\begin{aligned} \tilde{X}(t+h) = & x + \varepsilon \sum_{r=1}^q \sigma_r I_r + h(a + \varepsilon^2 b) + \varepsilon^2 \sum_{i,r=1}^q \Lambda_i \sigma_r I_{ir} + \varepsilon \sum_{r=1}^q (L_1 + \varepsilon^2 L_2) \sigma_r I_{or} \\ & + \varepsilon \sum_{r=1}^q \Lambda_r (a + \varepsilon^2 b) I_{ro} + h^2 (L_1 + \varepsilon^2 L_2)(a + \varepsilon^2 b) / 2 + h^3 L_1^2 a / 6. \end{aligned} \quad (3.8)$$

Using the properties of the Ito integral we obtain

$$\begin{aligned} |E\tilde{\rho}| &\leq K(x)[h^4 + \varepsilon^2 h^3], \quad E\tilde{\rho}^2 \leq K(x)[h^8 + \varepsilon^2 h^5 + \varepsilon^6 h^3], \quad |E(\tilde{\rho} \cdot \varepsilon I_r)| \leq K(x)\varepsilon^2 h^3, \\ |E(\tilde{\rho} \cdot \varepsilon^2 I_{ir})| &\leq K(x)\varepsilon^4 h^3, \quad |E(\tilde{\rho} \cdot \varepsilon I_{ro})| \leq K(x)\varepsilon^2 h^4, \\ |E(\tilde{\rho} \cdot \varepsilon I_{or})| &\leq K(x)\varepsilon^2 h^4, \quad |E(\tilde{\rho} \cdot \varepsilon I_i \cdot \varepsilon I_r)| \leq K(x)\varepsilon^4 h^3, \quad i, r = 1, \dots, q, \end{aligned} \quad (3.9)$$

which yields the inequalities

$$(3.10) \quad \left| E \prod_{j=1}^m \Delta^{i_j} - E \prod_{j=1}^m \tilde{\Delta}^{i_j} \right| \leq K(x)(h^4 + \varepsilon^2 h^3), \quad m = 1, \dots, 5,$$

$$\tilde{\Delta}^{i_j} = \tilde{X}^{i_j}(t+h) - x^{i_j}, \quad i_j = 1, \dots, n.$$

The inequalities (3.10) are valid because (1) the inequalities (3.10) for $m = 1, 2$ follow directly from (3.9); (2) it is clear that the odd (third and fifth) moments have a degree of smallness not worse than the first moments, while the fourth moments are not worse than the second.

The approximation \tilde{X} uses multiple Ito integrals I_{ir} which cannot be easily simulated. For this reason we shall try, as previously in [7], [8], to construct yet another approximation \bar{X} such that the inequalities

$$(3.11) \quad \left| E \prod_{j=1}^m \bar{\Delta}^{i_j} - E \prod_{j=1}^m \bar{\Delta}^{i_j} \right| \leq K(x)(h^4 + \varepsilon^2 h^3), \quad i_j = 1, \dots, n, \quad m = 1, \dots, 5$$

hold but \bar{X} contains only sufficiently simple random variables. Note that the inequalities (3.1) for $m = 1, \dots, 5$ and with right-hand sides $K(x)(h^4 + \varepsilon^2 h^3)$ follow directly from (3.10) and (3.11).

Let us now construct the approximation \bar{X} on the basis of \tilde{X} in the form

$$(3.12) \quad \begin{aligned} \bar{X} = & x + \varepsilon h^{1/2} \sum_{r=1}^q \sigma_r \xi_r + h(a + \varepsilon^2 b) + \varepsilon^2 h \sum_{i,r=1}^q \Lambda_i \sigma_r \xi_{ir} + \varepsilon h^{3/2} \sum_{r=1}^q (L_1 + \varepsilon^2 L_2) \sigma_r (\xi_r - \eta_r) \\ & + \varepsilon h^{3/2} \sum_{r=1}^q \Lambda_r (a + \varepsilon^2 b) \eta_r + h^2 (L_1 + \varepsilon^2 L_2) (a + \varepsilon^2 b) / 2 + h^3 L_1^2 a / 6, \end{aligned}$$

where the random variables ξ_r, ξ_{ir}, η_r are such that the inequalities (3.2) with $s = 6$ and with right-hand sides $K(x)(h^4 + \varepsilon^2 h^3)$, as well as the inequalities (3.11), hold. To satisfy these inequalities it is sufficient, for instance, to simulate the used random variables as in (3.7) (see [7, 8]). The inequalities (3.2) and (3.3) with $s = 6$ and with right-hand sides $K(x)(h^4 + \varepsilon^2 h^3)$ are obviously valid because each term of $\Delta = X - x$ and of $\bar{\Delta} = \bar{X} - x$ has a degree of smallness with respect to h which is at least $1/2$, while the term with order $1/2$ is multiplied by ε .

So we have proved the inequalities (3.1)–(3.3) with $s = 6$ and with right-hand sides $K(x)(h^4 + \varepsilon^2 h^3)$ for the approximation \bar{X} . If we now expand $f(X(t+h))$ in powers of $\Delta^i = X^i - x^i$ and $f(\bar{X}(t+h))$ in powers of $\bar{\Delta}^i = \bar{X}^i - x^i$ at the point x by means of the Taylor formula with Lagrangian remainder containing terms of order 6, then, as mentioned above, (3.1)–(3.3) imply the inequality (2.2) with right-hand sides $K(x)(h^4 + \varepsilon^2 h^3)$. Hence, according to Theorem 2.1, the weak method

$$(3.13) \quad \begin{aligned} X_{k+1} = & X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k + \varepsilon^2 h \sum_{i,r=1}^q (\Lambda_i \sigma_r \xi_{ir})_k \\ & + \varepsilon h^{3/2} \sum_{r=1}^q ((L_1 + \varepsilon^2 L_2) \sigma_r \xi_r)_k / 2 + \varepsilon h^{3/2} \sum_{r=1}^q (\Lambda_r (a + \varepsilon^2 b) \xi_r)_k / 2 \\ & + h^2 ((L_1 + \varepsilon^2 L_2) (a + \varepsilon^2 b))_k / 2 + h^3 (L_1^2 a)_k / 6 \end{aligned}$$

corresponding to the one-step approximation \bar{X} from (3.12), with the random variables $(\xi_r)_k$ and $(\xi_{ir})_k$ simulated at each step according to (3.7), has the global error $O(h^3 + \varepsilon^2 h^2)$. This method is of second order in h just like the standard second-order method (3.6), which provides an error $O(h^2)$. But the term h^2 in the error of the method (3.13) is multiplied by ε^2 . That is why the new method has smaller error than the standard scheme (3.6).

Above we shifted a term which is sufficiently simple to simulate from the remainder to the method, thereby reducing the error. However, we can also shift some more complicated terms, multiplied by ε^α , from the method to the corresponding remainder. Such a procedure reduces the computational costs (which, of course, is important for applications) while it does not lead to a substantial increase of the error.

For instance, by shifting the complicated (from a computational point of view) terms $\varepsilon^3 h^{3/2} L_2 \sigma_r \xi_r / 2$ and $\varepsilon^4 h^2 L_2 b / 2$ from the weak approximation (3.12) to its remainder we obtain a further method for solving the system (1.1). It can be seen that such a method has a global error $O(h^3 + \varepsilon^4 h)$. Moreover, if we additionally transfer the terms $\varepsilon^2 h \Lambda_i \sigma_r \xi_{ir}$ and $\varepsilon^3 h^{3/2} \Lambda_r b \xi_r / 2$ from approximation (3.12) to its remainder, it can be proved (such a proof essentially uses the equalities $E\xi_r = 0$, $E\xi_{ir} = 0$, $E\xi_{ir}\xi_j = 0$) that we do not lose the accuracy of the corresponding method, with respect to both h and ε . We finally arrive at the method

$$\begin{aligned} X_{k+1} = X_k &+ \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k + \varepsilon h^{3/2} \sum_{r=1}^q (L_1 \sigma_r \xi_r)_k / 2 \\ &+ \varepsilon h^{3/2} \sum_{r=1}^q (\Lambda_r a \xi_r)_k / 2 + h^2 ((L_1 + \varepsilon^2 L_2) a)_k / 2 \\ &+ \varepsilon^2 h^2 (L_1 b)_k / 2 + h^3 (L_1^2 a)_k / 6, \end{aligned} \quad (3.14)$$

the global error of which is $O(h^3 + \varepsilon^4 h)$. For its realization, it is sufficient to simulate only q independent random variables ξ_r according to the law $P(\xi = -1) = P(\xi = 1) = 1/2$. The time-step order of the method (3.14) is equal to 1, i.e., it is lower than the time-step order of the method (3.13) and of the standard method (3.6). Nevertheless, for small ε the error behavior of method (3.14) is acceptable. If we, for instance, choose a time-step h with $h = C\varepsilon^\alpha$, $0 < \alpha < 4$, method (3.14) even beats the standard method (3.6) as far as the degree of smallness with respect to ε is concerned. Furthermore, if we choose a time-step $h = C\varepsilon^\alpha$, $0 < \alpha \leq 2$, then method (3.14) is not worse than method (3.13) in the same sense. We want to emphasize that, additionally, method (3.14) requires fewer calculations of both the number of simulated random variables and the number of arithmetic operations.

Thus, we have briefly explained how to construct weak methods for a system with small noise. Let us stress that to put these new methods on a sound basis, one must thoroughly analyze the remainder, prove relations like (3.9) and inequalities like (3.1)–(3.3) for appropriate p , S , and $J(l)$, and then apply Theorem 2.1.

4. Taylor-type weak methods. We aim at constructing weak methods that have small errors (under the condition that ε is a small parameter) and are sufficiently effective with respect to their computational costs. For the system (1.1) we obtain Taylor-type weak methods with errors $O(h^2 + \varepsilon^2 h)$, $O(h^2 + \varepsilon^4 h)$, $O(h^3 + \varepsilon^2 h)$, $O(h^3 + \varepsilon^4 h)$, $O(h^3 + \varepsilon^2 h^2)$, $O(h^3 + \varepsilon^4 h^2)$, $O(h^4 + \varepsilon^2 h)$, $O(h^4 + \varepsilon^2 h^2 + \varepsilon^4 h)$, $O(h^4 + \varepsilon^4 h)$, $O(h^4 + \varepsilon^2 h^2)$, $O(h^4 + \varepsilon^4 h^2)$. In section 3 we have derived methods with errors $O(h^3 + \varepsilon^4 h)$ and $O(h^3 + \varepsilon^2 h^2)$. More methods can be derived in the same manner. In this

section we state several methods with errors $O(h^4 + \dots)$. Others can be found in the preprint [11]. There one also can find some implicit methods.

By the approach stated above, it is possible to derive methods with errors $O(h^5 + \dots)$, $O(h^6 + \dots)$, etc. But we do not write them down because most popular deterministic schemes have orders not higher than 4. Note that it is also possible to derive methods with errors $O(h^3 + \varepsilon^6 h^2)$, $O(h^4 + \varepsilon^6 h^2)$, $O(h^4 + \varepsilon^\alpha h^3)$, $\alpha = 2, \dots, 8$, but they require huge computational efforts.

We introduce the notation R for the global error of a method.

Note that all random variables used are mutually independent.

Methods $O(h^4 + \dots)$. Our method with error $R = O(h^4 + \varepsilon^2 h)$ has the form

$$(4.1) \quad \begin{aligned} X_{k+1} = & X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k \\ & + h^2 (L_1 a)_k / 2 + h^3 (L_1^2 a)_k / 6 + h^4 (L_1^3 a)_k / 24, \end{aligned}$$

where the random variables ξ_r are simulated using

$$(4.2) \quad P(\xi = -1) = P(\xi = 1) = \frac{1}{2}.$$

Our method with error $R = O(h^4 + \varepsilon^4 h)$ has the form

$$(4.3) \quad \begin{aligned} X_{k+1} = & X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k + \varepsilon h^{3/2} \sum_{r=1}^q (L_1 \sigma_r (\xi_r / 2 - \eta_r))_k \\ & + \varepsilon h^{3/2} \sum_{r=1}^q (\Lambda_r a (\xi_r / 2 + \eta_r))_k + h^2 (L_1 (a + \varepsilon^2 b))_k / 2 + \varepsilon^2 h^2 (L_2 a)_k / 2 \\ & + \varepsilon h^{5/2} \sum_{r=1}^q ((L_1^2 \sigma_r + L_1 \Lambda_r a + \Lambda_r L_1 a) \xi_r)_k / 6 + h^3 (L_1^2 (a + \varepsilon^2 b))_k / 6 \\ & + \varepsilon^2 h^3 ((L_1 L_2 + L_2 L_1) a)_k / 6 + h^4 (L_1^3 a)_k / 24, \end{aligned}$$

where the random variables ξ_r and η_r are simulated as

$$(4.4) \quad P(\xi = -1) = P(\xi = 1) = \frac{1}{2}, \quad P\left(\eta = -\frac{1}{\sqrt{12}}\right) = P\left(\eta = \frac{1}{\sqrt{12}}\right) = \frac{1}{2}.$$

Our method with error $R = O(h^4 + \varepsilon^2 h^2)$ has the form

$$(4.5) \quad \begin{aligned} X_{k+1} = & X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k + \varepsilon^2 h \sum_{i,r=1}^q (\Lambda_i \sigma_r \xi_{ir})_k \\ & + \varepsilon h^{3/2} \sum_{r=1}^q ((L_1 + \varepsilon^2 L_2) \sigma_r \xi_r)_k / 2 + \varepsilon h^{3/2} \sum_{r=1}^q (\Lambda_r (a + \varepsilon^2 b) \xi_r)_k / 2 \\ & + h^2 ((L_1 + \varepsilon^2 L_2) (a + \varepsilon^2 b))_k / 2 + h^3 (L_1^2 a)_k / 6 + h^4 (L_1^3 a)_k / 24, \end{aligned}$$

where the random variables ξ_r and ξ_{ir} are simulated according to [7, 8]:

$$P(\xi = 0) = \frac{2}{3}, \quad P(\xi = -\sqrt{3}) = P(\xi = \sqrt{3}) = \frac{1}{6}, \quad \xi_{ir} = (\xi_i \xi_r - \gamma_{ir} \zeta_i \zeta_r) / 2,$$

$$(4.6) \quad \gamma_{ir} = \begin{cases} -1, & i < r \\ 1, & i \geq r \end{cases}, \quad P(\zeta = -1) = P(\zeta = 1) = \frac{1}{2}$$

or [14]:

$$(4.7) \quad \begin{aligned} P(\xi = 0) &= \frac{2}{3}, & P(\xi = -\sqrt{3}) &= P(\xi = \sqrt{3}) = \frac{1}{6}, \\ \xi_{ir} &= (\xi_i \xi_r - \zeta_{ir})/2, & \zeta_{ii} &= 1, & \zeta_{ir} &= -\zeta_{ri}, & i \neq r, \\ P(\zeta_{ir} = -1) &= P(\zeta_{ir} = 1) = \frac{1}{2}, & i < r. \end{aligned}$$

Our method with error $R = O(h^4 + \varepsilon^4 h^2)$ has the form

$$(4.8) \quad \begin{aligned} X_{k+1} &= X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k + \varepsilon^2 h \sum_{i,r=1}^q (\Lambda_i \sigma_r \xi_{ir})_k \\ &\quad + \varepsilon h^{3/2} \sum_{r=1}^q ((L_1 + \varepsilon^2 L_2) \sigma_r (\xi_r - \mu_r))_k + \varepsilon h^{3/2} \sum_{r=1}^q (\Lambda_r (a + \varepsilon^2 b) \mu_r)_k \\ &\quad + h^2 ((L_1 + \varepsilon^2 L_2)(a + \varepsilon^2 b))_k / 2 + \varepsilon h^{5/2} \sum_{r=1}^q ((L_1^2 \sigma_r + L_1 \Lambda_r a + \Lambda_r L_1 a) \xi_r)_k / 6 \\ &\quad + h^3 (L_1^2 (a + \varepsilon^2 b))_k / 6 + \varepsilon^2 h^3 ((L_1 L_2 + L_2 L_1) a)_k / 6 + h^4 (L_1^3 a)_k / 24, \end{aligned}$$

where ξ_r , ξ_{ir} , and μ_r are simulated, for example, according to

$$(4.9) \quad \begin{aligned} \xi_{ir} &= (\xi_i \xi_r - \gamma_{ir} \zeta_i \zeta_r) / 2, & \gamma_{ir} &= \begin{cases} -1, & i < r \\ 1, & i \geq r \end{cases}, & P(\zeta = -1) &= P(\zeta = 1) = \frac{1}{2}, \\ P(\xi = 0) &= \frac{2}{3}, & P(\xi = -\sqrt{3}) &= P(\xi = \sqrt{3}) = \frac{1}{6}, & \mu_r &= \xi_r / 2 + \zeta_r / \sqrt{12}. \end{aligned}$$

Remark on how to choose the increment h given ε . Below we discuss the interdependence of the time increment h and the parameter ε in the methods of this section.

We first choose the time increment h to be $h = C\varepsilon^\alpha$. Then the global error of a method can be estimated in powers of the small parameter ε by

$$R = O(\varepsilon^\beta),$$

where

$$\beta = \min \left\{ \alpha p, \min_{l \in S} (\alpha l + J(l)) \right\}.$$

If $h = C\varepsilon^\alpha$, method (4.8) has $R = O(\varepsilon^{4\alpha} + \varepsilon^{2\alpha+4})$, while the method (4.1) yields $R = O(\varepsilon^{4\alpha} + \varepsilon^{\alpha+2})$. In the case of $0 < \alpha \leq 2/3$, both errors are bounded by $O(\varepsilon^{4\alpha})$, and so both methods have the same order with respect to ε . However, if $\alpha > 2/3$, method (4.8) has higher order with respect to ε than (4.1) (for instance, if $\alpha = 2$, we have $O(\varepsilon^8)$ for (4.8) and $O(\varepsilon^4)$ for (4.1)). Thus, in the case of a comparatively large time increment h compared to ε (this is of interest mainly if ε is sufficiently small, i.e., when the error, estimated by ε^β , is not large), complicated methods like (4.8) and sufficiently simple methods like (4.1) have the same order in ε . In such a situation simple methods are usually preferable because of their considerably lower computational costs. But if one wants to reach an error of high order with respect to ε , complicated methods are preferable.

5. Weak Runge–Kutta methods. To avoid costly calculations of derivatives in the methods of section 4 we propose Runge–Kutta schemes. Below we consider (i) full (derivative-free) Runge–Kutta schemes and (ii) Runge–Kutta schemes without derivatives of the coefficients $a(t, x)$ and $b(t, x)$ but with derivatives of the diffusion coefficients $\sigma_r(t, x)$ (semi-Runge–Kutta schemes) which may be useful in the case of simple functions σ_r .

It is known (see [5], [8]) that in the case of a general system with $\varepsilon \equiv 1$ there are no entirely constructive higher-order Runge–Kutta schemes. For systems with small noises we obtain full Runge–Kutta methods with errors $O(h^2 + \varepsilon^2 h)$, $O(h^2 + \varepsilon^4 h)$, $O(h^3 + \varepsilon^2 h)$, $O(h^3 + \varepsilon^4 h)$, $O(h^4 + \varepsilon^2 h)$, $O(h^4 + \varepsilon^2 h^2 + \varepsilon^4 h)$, and $O(h^4 + \varepsilon^4 h)$. For higher orders we have succeeded in constructing semi-Runge–Kutta schemes with errors $O(h^3 + \varepsilon^2 h^2)$, $O(h^3 + \varepsilon^4 h^2)$, $O(h^4 + \varepsilon^2 h^2)$ and $O(h^4 + \varepsilon^4 h^2)$.

In this section we state several methods with errors $O(h^2 + \dots)$ and $O(h^4 + \dots)$. Other Runge–Kutta methods for general systems with small noise and full Runge–Kutta methods for systems with small additive noise (for instance, full Runge–Kutta method with the error $O(h^4 + \varepsilon^4 h^2)$) can be found in our preprint [11]. Note that in [11] we also give a few full Runge–Kutta methods for systems with small colored noise, for instance, a scheme with error $O(h^4 + \varepsilon^2 h^3)$ and a full Runge–Kutta method with error $O(h^3)$ for general systems with colored noise, which has not been proposed previously (see [10] and references therein).

In the appendix (section 10) we give a detailed derivation of our semi-Runge–Kutta scheme $O(h^4 + \varepsilon^2 h^2)$. The other Runge–Kutta methods can be derived in the same manner.

To construct Runge–Kutta methods for system (1.1) we use as a subsidiary tool deterministic Runge–Kutta methods. To this end we select specific deterministic schemes which from our point of view are most appropriate. Obviously, it is possible to derive families of stochastic Runge–Kutta methods which are similar to the proposed ones but use different deterministic Runge–Kutta schemes.

Methods $O(h^2 + \dots)$. Our method with error $R = O(h^2 + \varepsilon^2 h)$ has the form

$$(5.1) \quad X_{k+1} = X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + \varepsilon^2 h b_k + h(a_k + a(t_{k+1}, X_k + h a_k))/2,$$

where ξ_r are as in (4.2).

Our method with error $R = O(h^2 + \varepsilon^4 h)$ has the form

$$(5.2) \quad X_{k+1} = X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r(t_k, X_k) + \sigma_r(t_{k+1}, X_k + h a_k)) \xi_{r_k} / 2 + h \left(a_k + a(t_{k+1}, X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k) \right) / 2 + \varepsilon^2 h (b_k + b(t_{k+1}, X_k + h a_k)) / 2,$$

where ξ_r are as in (4.2).

Methods $O(h^4 + \dots)$. Our method with error $R = O(h^4 + \varepsilon^2 h)$ has the form

$$(5.3) \quad X_{k+1} = X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + \varepsilon^2 h b_k + (k_1 + 2k_2 + 2k_3 + k_4)/6,$$

where

$$k_1 = h a_k, \quad k_2 = h a(t_{k+1/2}, X_k + k_1/2), \quad k_3 = h a(t_{k+1/2}, X_k + k_2/2),$$

$$(5.4) \quad k_4 = ha(t_{k+1}, X_k + k_3),$$

and ξ_r are as in (4.2).

Our method with error $R = O(h^4 + \varepsilon^2 h^2 + \varepsilon^4 h)$ has the form

$$(5.5) \quad \begin{aligned} X_{k+1} = X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r(t_k, X_k) + \sigma_r(t_{k+1}, X_k + ha_k)) \xi_{rk} / 2 \\ + (k_1 + 2k_2 + 2k_3 + k_4) / 6 + \varepsilon^2 h (b_k + b(t_{k+1}, X_k + ha_k)) / 2, \end{aligned}$$

where

$$(5.6) \quad \begin{aligned} k_1 = ha_k, \quad k_2 = ha(t_{k+1/2}, X_k + k_1/2), \quad k_3 = ha \left(t_{k+1/2}, X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + k_2/2 \right), \\ k_4 = ha \left(t_{k+1}, X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + k_3 + 3\varepsilon^2 h b_k \right), \end{aligned}$$

and ξ_r are as in (4.2).

Our method with error $R = O(h^4 + \varepsilon^4 h)$ has the form

$$(5.7) \quad \begin{aligned} X_{k+1} = X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r(t_k, X_k) (\xi_r + 6\eta_r)_k + 4\sigma_r(t_{k+1/2}, X_k + k_2/2) \xi_{rk} \\ + \sigma_r(t_{k+1}, X_k + k_1) (\xi_r - 6\eta_r)_k) / 6 + h \left(a \left(t_k, X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \eta_r)_k \right) \right. \\ \left. - a \left(t_k, X_k - \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \eta_r)_k \right) \right) / 2 + (k_1 + 2k_2 + 2k_3 + k_4) / 6 + \varepsilon^2 (l_1 + 3l_2) / 4, \end{aligned}$$

where

$$(5.8) \quad \begin{aligned} k_1 = ha_k, \quad k_2 = ha(t_{k+1/2}, X_k + k_1/2), \\ k_3 = ha \left(t_{k+1/2}, X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + k_2/2 + \varepsilon^2 l_1/4 + 3\varepsilon^2 l_2/4 \right), \\ k_4 = ha \left(t_{k+1}, X_k + \varepsilon h^{1/2} \sum_{r=1}^q \sigma_r(t_{k+1}, X_k + k_1) \xi_{rk} + k_3 + \varepsilon^2 l_1 \right), \\ l_1 = hb_k, \quad l_2 = hb(t_k + 2h/3, X_k + 2k_1/9 + 4k_2/9), \end{aligned}$$

and ξ_r, η_r are simulated as in (4.4). This full Runge–Kutta method requires six recalculations of the function $a(t, x)$, three recalculations of the functions $\sigma_r(t, x)$, and two recalculations of the function $b(t, x)$.

Our method with error $R = O(h^4 + \varepsilon^2 h^2)$ has the form

$$X_{k+1} = X_k + \varepsilon h^{1/2} \sum_{r=1}^q \left(\sigma_r(t_k, X_k) + \sigma_r(t_{k+1}, X_k) \right) \xi_{rk} / 2 + \varepsilon^2 h \sum_{i,r=1}^q (\Lambda_i \sigma_r \xi_{ir})_k$$

$$\begin{aligned}
& +\varepsilon h^{3/2} \sum_{r=1}^q \sum_{i=1}^n \left(a^i \frac{\partial \sigma_r}{\partial x^i} \xi_r \right)_k / 2 + \varepsilon^3 h^{3/2} \sum_{r=1}^q (L_2 \sigma_r \xi_r)_k / 2 + (k_1 + 2k_2 + 2k_3 + k_4)/6 \\
(5.9) \quad & + \varepsilon^2 h \left(b_k + b \left(t_{k+1}, X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k \right) \right) / 2,
\end{aligned}$$

where k_i , $i = 1, \dots, 4$, are from (5.6) and the used random variables ξ_r , ξ_{ir} are simulated as in method (4.5). Method (5.9) contains first and second derivatives of the functions σ_r with respect to x . This method is derived in the appendix (see section 10).

Note that for a single noise term ($q = 1$) we succeeded in constructing a full Runge–Kutta method with error $O(h^4 + \varepsilon^2 h^2)$ (see [11]).

Our method with error $R = O(h^4 + \varepsilon^4 h^2)$ has the form

$$\begin{aligned}
X_{k+1} = & X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r(t_k, X_k)(\xi_r + 6\eta_r)_k + 4\sigma_r(t_{k+1/2}, X_k + k_2/2)\xi_{rk} \\
& + \sigma_r(t_{k+1}, X_k + k_1)(\xi_r - 6\eta_r)_k)/6 + h \left(a \left(t_k, X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \eta_r)_k \right) \right. \\
& \left. - a \left(t_k, X_k - \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \eta_r)_k \right) \right) / 2 + \varepsilon^2 h \sum_{i,r=1}^q (\Lambda_i \sigma_r \xi_{ir})_k \\
(5.10) \quad & + \varepsilon^3 h^{3/2} \sum_{r=1}^q (L_2 \sigma_r \xi_r)_k / 2 + (k_1 + 2k_2 + 2k_3 + k_4)/6 + \varepsilon^2 (l_1 + 3l_2)/4,
\end{aligned}$$

where

$$\begin{aligned}
k_1 &= ha_k, & k_2 &= ha(t_{k+1/2}, X_k + k_1/2), \\
k_3 &= ha \left(t_{k+1/2}, X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + k_2/2 + \varepsilon^2 l_1/4 + 3\varepsilon^2 l_2/4 \right), \\
k_4 &= ha \left(t_{k+1}, X_k + \varepsilon h^{1/2} \sum_{r=1}^q \sigma_r(t_{k+1}, X_k + k_1)\xi_{rk} + k_3 + \varepsilon^2 l_1 \right), \\
l_1 &= hb \left(t_k, X_k + \varepsilon h^{1/2}(1 + \sqrt{3}) \sum_{r=1}^q (\sigma_r \xi_r)_k / 2 \right), & l_2 &= hb \left(t_k + 2h/3, X_k + 2\varepsilon^2 l_1/3 \right. \\
(5.11) \quad & \left. + 2k_1/9 + 4k_2/9 + \varepsilon h^{1/2}(3 - \sqrt{3}) \sum_{r=1}^q (\sigma_r \xi_r)_k / 6 \right),
\end{aligned}$$

and the used random variables are simulated using

$$\begin{aligned}
P(\xi = 0) &= \frac{2}{3}, & P(\xi = -\sqrt{3}) &= P(\xi = \sqrt{3}) = \frac{1}{6}, & P(\zeta = -1) &= P(\zeta = 1) = \frac{1}{2}, \\
(5.12) \quad \xi_{ir} &= (\xi_i \xi_r - \gamma_{ir} \zeta_i \zeta_r)/2, & \gamma_{ir} &= \begin{cases} -1, & i < r, \\ 1, & i \geq r, \end{cases} & \eta_r &= \zeta_r / \sqrt{12}.
\end{aligned}$$

6. Stratonovich system with small noise. It is known that a stochastic system in the Stratonovich sense (marked by “*”),

$$(6.1) \quad dX = a(t, X)dt + \varepsilon^2 c(t, X)dt + \varepsilon \sum_{r=1}^q \sigma_r(t, X) * dW_r, \quad X(t_o) = X_o, \quad t \in [t_o, T],$$

is equivalent to the following system in the Ito sense:

$$(6.2) \quad dX = a(t, X)dt + \varepsilon^2 b(t, X)dt + \varepsilon \sum_{r=1}^q \sigma_r(t, X)dW_r,$$

where

$$(6.3) \quad b(t, x) = c(t, x) + \frac{1}{2} \sum_{r=1}^q \frac{\partial \sigma_r}{\partial x}(t, x) \sigma_r(t, x).$$

In sections 4 and 5 we have proposed weak methods for the Ito system having the form of (6.2). Thus, the methods of sections 4 and 5 are also appropriate for the Stratonovich system (6.1). Note that the full Runge–Kutta methods of section 5 are no longer full when applied to system (6.1), since $b(t, x)$ in (6.3) contains derivatives $\frac{\partial \sigma_r}{\partial x}$. However, if the diffusion coefficients σ_r are simple functions, the methods of section 5 may be efficient and useful for the Stratonovich system (6.1). Nevertheless, in some cases we obtain the full Runge–Kutta schemes for (6.1) (see [11]).

7. Talay–Tubaro expansion of the global error. Talay and Tubaro proved in [16] that it is possible to expand the global errors of methods for stochastic systems in powers of time increment h . Their approach is analogous to the Runge extrapolation method for ordinary differential equations and allows us to estimate the global error as well as to improve the accuracy of the method. Below we expand the global error not only in powers of the time increment h , but also in powers of the small parameter ε . Therefore, we cannot directly apply the Talay–Tubaro theorem.

THEOREM 7.1. *The global error of the method*

$$(7.1) \quad X_{k+1} = X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k + h^2(L_1 a)_k/2,$$

$$P(\xi = -1) = P(\xi = 1) = \frac{1}{2}$$

is

$$(7.2) \quad R = O(h^2 + \varepsilon^2 h) = C_1(\varepsilon)h^2 + \varepsilon^2 C_2(\varepsilon)h + O(h^3 + \varepsilon^2 h^2),$$

where the functions $C_i(\varepsilon)$, $i = 1, 2$, do not depend on h and are equal to $C_i(\varepsilon) = C_i^o + O(\varepsilon^2)$, and the constants C_i^o , $i = 1, 2$, do not depend on both h and ε .

For the proof of Theorem 7.1 and the expressions for the coefficients $C_i(\varepsilon)$, we refer to the appendix (section 11). The same proof shows that the expansions of the global error for other methods can be obtained in the same way as expansion (7.2) for method (7.1). For instance, for method (4.3) with error $O(h^4 + \varepsilon^4 h)$, we have

$$R = C_1(\varepsilon)h^4 + \varepsilon^2 C_2(\varepsilon)h^3 + \varepsilon^4 C_3(\varepsilon)h^2 + \varepsilon^4 C_4(\varepsilon)h + O(h^5 + \varepsilon^6 h^2).$$

An expansion like (7.2) can be used to derive extrapolation schemes as follows. Simulate $u^\varepsilon(t_o, X_o) = Ef(X_{t_o, X_o}^\varepsilon(T))$ twice using method (7.1) for given ε but with varying time-steps $h_1 = h$, $h_2 = \alpha h$, $\alpha > 0$, $\alpha \neq 1$. We obtain $\bar{u}^{\varepsilon, h_1}(t_o, X_o) = Ef(\bar{X}_{t_o, X_o}^{\varepsilon, h_1}(T))$, $\bar{u}^{\varepsilon, h_2}(t_o, X_o) = Ef(\bar{X}_{t_o, X_o}^{\varepsilon, h_2}(T))$. We can expand

$$u^\varepsilon = \bar{u}^{\varepsilon, h_1} + C_1(\varepsilon)h_1^2 + \varepsilon^2 C_2(\varepsilon)h_1 + O(h^3 + \varepsilon^2 h^2)$$

and

$$u^\varepsilon = \bar{u}^{\varepsilon, h_2} + C_1(\varepsilon)h_2^2 + \varepsilon^2 C_2(\varepsilon)h_2 + O(h^3 + \varepsilon^2 h^2).$$

This yields

$$(7.3) \quad \varepsilon^2 C_2(\varepsilon) = \varepsilon^2 \bar{C}_2(\varepsilon) - C_1^o \cdot (h_1 + h_2) + O(h^2 + \varepsilon^2 h),$$

where $\varepsilon^2 \bar{C}_2(\varepsilon)$ is given by

$$\varepsilon^2 \bar{C}_2(\varepsilon) = (\bar{u}^{\varepsilon, h_1} - \bar{u}^{\varepsilon, h_2}) / (h_2 - h_1).$$

On the other hand, using method (7.1) and setting $\varepsilon = 0$ with varying time-steps h_1, h_2 , we obtain $\bar{u}^{o, h_1}(t_o, X_o) = f(\bar{X}_{t_o, X_o}^{o, h_1}(T))$, $\bar{u}^{o, h_2}(t_o, X_o) = f(\bar{X}_{t_o, X_o}^{o, h_2}(T))$, where $\bar{X}_{t_o, X_o}^{o, h_i}(t)$ is the corresponding approximation to the solution $X_{t_o, X_o}^o(t)$ of the deterministic system. Then the Runge extrapolation method yields

$$(7.4) \quad C_1(0) = C_1^o = \bar{C}_1^o + O(h),$$

where \bar{C}_1^o can be calculated by

$$\bar{C}_1^o = (\bar{u}^{o, h_1} - \bar{u}^{o, h_2}) / (h_2^2 - h_1^2).$$

By (7.3) and (7.4) we obtain an improved value $\bar{u}_{imp}^\varepsilon$ with error $O(h^3 + \varepsilon^2 h^2)$ by letting

$$(7.5) \quad \bar{u}_{imp}^\varepsilon = \bar{u}^{\varepsilon, h_1} + \varepsilon^2 \bar{C}_2(\varepsilon)h_1 - \bar{C}_1^o h_1 h_2.$$

In the same spirit, using three recalculations of $u^\varepsilon(t_o, X_o) = Ef(X_{t_o, X_o}^\varepsilon(T))$ by method (7.1) for given ε and with varying time-steps, one can also find $C_1(\varepsilon)$ and $C_2(\varepsilon)$ from (7.2) and obtain yet another improved value.

We conclude that according to our approach to the construction of weak methods for a system with small noise, we can shift some terms, which contribute to the error proportionally to $h^i \varepsilon^j$, from the method to its remainder and vice versa. By calculating the constants $C_i(\varepsilon)$ it is possible to estimate the proper weights of the terms in the sums above and to select the most appropriate scheme for solving a given system with small noise, keeping both computational costs low and accuracy high.

8. Reduction of the Monte-Carlo error. Let us calculate the expectation $Ef(X(T))$ by a Monte-Carlo technique using a weak method for the solution of system (1.1). Then, as is already known, two kinds of errors arise: the error of the weak method, which was considered in the previous sections, and the Monte-Carlo error which will be discussed below. We have

$$(8.1) \quad Ef(\bar{X}(T)) \approx \frac{1}{N} \sum_{m=1}^N f(\bar{X}^{(m)}(T)) \pm \frac{c}{\sqrt{N}} [Df(\bar{X}(T))]^{1/2},$$

where N is the number of independent realizations $\bar{X}^{(m)}$ simulated by a weak method and c is some constant. If, for instance, the constant c is equal to 1, 2, or 3, the corresponding calculated value belongs to the interval defined by (8.1) with confidence probability 0.68, 0.95, or 0.997, respectively.

If $Df(\bar{X}(T))$ is close to $Df(X(T))$, the Monte-Carlo error can be estimated by $c[Df(X(T))/N]^{1/2}$. If $Df(X(T))$ is large, to achieve the required accuracy, we must choose N sufficiently large, which increases the computational costs. If we succeed in constructing a variable Z instead of $f(X(T))$, such that $EZ = Ef(X(T))$ but $DZ \ll Df(X(T))$, a simulation of the variable Z instead of $f(X(T))$ would yield more accurate results without increasing the computational effort.

One approach to constructing such a Z was proposed in [8]. As will be shown below, this approach allows us to effectively reduce the Monte-Carlo error of a system with small noise.

Together with system (1.1), consider the following:

$$\begin{aligned} dX &= a(t, X)dt + \varepsilon^2 b(t, X)dt - \varepsilon \sum_{r=1}^q \mu_r(t, X) \sigma_r(t, X)dt + \varepsilon \sum_{r=1}^q \sigma_r(t, X) dW_r, \\ (8.2) \quad dY &= \varepsilon \sum_{r=1}^q \mu_r(t, X) Y dW_r, \end{aligned}$$

where μ_r and Y are scalars.

According to the Girsanov theorem for any μ_r , we have

$$(8.3) \quad yEf(X_{s,x}(T))|_{(1.1)} = E(Y_{s,x,y}(T)f(X_{s,x}(T)))|_{(8.2)}.$$

The function $u(s, x) = Ef(X_{s,x}(T))|_{(1.1)}$ satisfies the equation

$$(8.4) \quad Lu \equiv \frac{\partial u}{\partial s} + \sum_{i=1}^n a^i \frac{\partial u}{\partial x^i} + \varepsilon^2 \sum_{i=1}^n b^i \frac{\partial u}{\partial x^i} + \frac{\varepsilon^2}{2} \sum_{r=1}^q \sum_{i=1}^n \sum_{j=1}^n \sigma_r^i \sigma_r^j \frac{\partial^2 u}{\partial x^i \partial x^j} = 0$$

subject to the following condition at the instant T :

$$(8.5) \quad u(T, x) = f(x).$$

Under sufficiently mild conditions on the coefficients and on the function f , the solution $u(s, x) = u^\varepsilon(s, x)$ of problem (8.4)–(8.5) has the form (see [3, Chap. 2])

$$(8.6) \quad u^\varepsilon(s, x) = u^o(s, x) + \varepsilon^2 u^1(s, x; \varepsilon).$$

The function u^o satisfies the first-order partial differential equation

$$(8.7) \quad \frac{\partial u}{\partial s} + \sum_{i=1}^n a^i \frac{\partial u}{\partial x^i} = 0$$

under condition (8.5). Obviously, the solution of (8.7) has the form

$$(8.8) \quad u^o(s, x) = f(X_{s,x}^o(T)),$$

where $X_{s,x}^o$ is the solution of the Cauchy problem for the deterministic system of differential equations

$$(8.9) \quad \frac{dX}{dt} = a(t, X), \quad X(s) = x.$$

Applying the Ito formula along the solution of the system (8.2) we get the following expression (note that here $Lu = 0$):

$$\begin{aligned} d[u(t, X_{s,x}(t))Y_{s,x,y}(t)] &= LuYdt - \varepsilon \sum_{r=1}^q \mu_r \left(\sigma_r, \frac{\partial u}{\partial x} \right) Y dt + \varepsilon \sum_{r=1}^q \left(\sigma_r, \frac{\partial u}{\partial x} \right) Y dW_r(t) \\ &+ u \sum_{r=1}^q \mu_r Y dW_r(t) + \varepsilon \sum_{r=1}^q \left(\sigma_r, \frac{\partial u}{\partial x} \right) \mu_r Y dt = \sum_{r=1}^q \left(\varepsilon \left(\sigma_r, \frac{\partial u}{\partial x} \right) + \mu_r u \right) Y dW_r(t). \end{aligned}$$

Then

$$(8.10) \quad u(t, X_{s,x}(t))Y_{s,x,y}(t) = u(s, x)y + \int_s^t \sum_{r=1}^q \left(\varepsilon \left(\sigma_r, \frac{\partial u}{\partial x} \right) + \mu_r u \right) Y dW_r(t).$$

If we suppose that $t = T$, $y = 1$, $\mu_r \equiv 0$, we obtain

$$f(X_{s,x}(T)) = u(s, x) + \int_s^T \varepsilon \sum_{r=1}^q \left(\sigma_r, \frac{\partial u}{\partial x} \right) dW_r(t).$$

Therefore

$$(8.11) \quad Df(X_{s,x}(T)) = \varepsilon^2 \int_s^T E \left[\sum_{r=1}^q \left(\sigma_r, \frac{\partial u}{\partial x} \right) \right]^2 dt$$

because $u(s, x) = Ef(X_{s,x}(T))|_{(1.1)}$.

Thus, if we calculate $Ef(X(T))$ by a Monte-Carlo technique using a weak method for solving system (1.1), then the Monte-Carlo error, evaluated by $c[Df(\bar{X}(T))/N]^{1/2}$ and close to $c[Df(X(T))/N]^{1/2}$, contains a small factor equal to ε .

As can be seen from (8.3), the mean value $EZ = E(Y_{s,x,y}(T)f(X_{s,x}(T)))|_{(8.2)}$ does not depend on μ_r , whereas $D(Y_{s,x,y}(T)f(X_{s,x}(T)))|_{(8.2)}$ does depend on μ_r . So below we will select functions μ_r , $r = 1, \dots, q$, such that the variance DZ becomes less than the variance (8.11).

Assume that $f > 0$. Then $u^o > 0$. Note that if the function f is not positive but there are constants K and C such that $Kf + C > 0$, we can take the function $g = Kf + C$ instead. Then we can simulate Eg and finally obtain Ef .

Setting $t = T$, $y = 1$ in formula (8.10) and

$$(8.12) \quad \mu_r = -\frac{\varepsilon}{u^o} \left(\sigma_r, \frac{\partial u^o}{\partial x} \right), \quad r = 1, \dots, q,$$

we obtain

$$f(X_{s,x}(T))Y = u(s, x) + \int_s^T \varepsilon^3 \sum_{r=1}^q \left[\left(\sigma_r, \frac{\partial u^1}{\partial x} \right) - \left(\sigma_r, \frac{\partial u^o}{\partial x} \right) \frac{u^1}{u^o} \right] dW_r(t).$$

Therefore

$$D[f(X_{s,x}(T))Y] = \varepsilon^6 \int_s^T E \left(\sum_{r=1}^q \left[\left(\sigma_r, \frac{\partial u^1}{\partial x} \right) - \left(\sigma_r, \frac{\partial u^o}{\partial x} \right) \frac{u^1}{u^o} \right] \right)^2 dt.$$

Hence, the Monte-Carlo error for system (8.2) with μ_r chosen from (8.12) inherits a small factor equal to ε^3 .

System (8.2) with μ_r from (8.12) is again a system with small noise, and all methods proposed above are suitable for finding its solution. We observe that, even if the number N of simulations is small, the Monte-Carlo error for this system will be reasonably small. Of course, in order to apply the approach outlined above, we must know the function $u^o(s, x)$.

9. Simulation of the Lyapunov exponent of a linear system with small noise. Lyapunov exponents are useful when investigating stability of a dynamic stochastic system [1], [4]. Talay [15] (see also [5]) previously proposed a numerical approach to calculating Lyapunov exponents, based on the ergodic property: using weak methods Lyapunov exponents are calculated by simulating a single trajectory. This procedure is appealing because it is intuitive and computationally cheap. However, it is difficult to analyze the errors arising from this approach.

Below we calculate the Lyapunov exponent as a convenient example to illustrate the correctness and efficiency of our proposed methods. We also pay attention to analysis of the errors.

For our numerical tests we take the following two-dimensional linear Ito stochastic system:

$$(9.1) \quad dX = AXdt + \varepsilon \sum_{r=1}^q B_r X dW_r,$$

where X is a two-dimensional vector, A and B_r are constant 2×2 -matrices, W_r are independent standard Wiener processes, and $\varepsilon > 0$ is a small parameter.

In the ergodic case there exists a unique Lyapunov exponent λ of system (9.1) (cf. [4]), with

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} E \rho(t) = \lim_{t \rightarrow \infty} \frac{1}{t} \rho(t) \quad \text{a.s.},$$

where $\rho(t) = \ln |X(t)|$, and $X(t)$, $t \geq 0$, is a nontrivial solution to system (9.1).

If $D(\rho(t)) \rightarrow 0$ for $t \rightarrow \infty$, then [4]

$$(9.2) \quad E \left(\frac{\rho(t)}{t} - \lambda \right)^2 = D \left(\frac{\rho(t)}{t} \right) (1 + \varphi^2(t)),$$

where $\varphi(t) \rightarrow 0$ for $t \rightarrow \infty$. It is not difficult to show that $D(\rho(t)/t) \rightarrow 0$ for $t \rightarrow \infty$. From (9.2) and the equality

$$D \left(\frac{\rho(t)}{t} \right) = E \left(\frac{\rho(t)}{t} - \lambda \right)^2 - \left[E \left(\frac{\rho(t)}{t} \right) - \lambda \right]^2,$$

we have

$$(9.3) \quad \left| E \left(\frac{\rho(t)}{t} \right) - \lambda \right| = \varphi(t) \left[D \left(\frac{\rho(t)}{t} \right) \right]^{1/2}.$$

Herein we consider system (9.1) with the matrices A and B_r of the form

$$(9.4) \quad A = \begin{pmatrix} a & c \\ -c & a \end{pmatrix}, \quad B_r = \begin{pmatrix} b_r & d_r \\ -d_r & b_r \end{pmatrix}, \quad r = 1, 2.$$

In this case the Lyapunov exponent is (cf. [2])

$$(9.5) \quad \lambda = a + \frac{\varepsilon^2}{2} \sum_{r=1}^2 [(d_r)^2 - (b_r)^2].$$

By a Monte-Carlo technique we numerically calculate the function

$$(9.6) \quad \lambda(T) = \frac{1}{T} E \rho(T) \approx \bar{\lambda}(T) = \frac{1}{T} E \bar{\rho}(T), \quad \bar{\rho}(T) = \ln |\bar{X}(T)|.$$

The function $\lambda(t)$ in the limit of large time ($t \rightarrow \infty$) tends to the Lyapunov exponent λ . In this case three errors arise: (1) the method error, i.e., $|E \rho(T)/T - E \bar{\rho}(T)/T|$, (2) the Monte-Carlo error which is bounded by $c[D(\bar{\rho}(T)/T)]^{1/2}/\sqrt{N}$ (see (8.1)), and (3) the error with respect to the choice of integration time T (see (9.3)).

As can be seen from our computational results, the third error, i.e., $|\lambda(T) - \lambda| = |E(\rho(T)/T) - \lambda|$, is negligibly small, at any rate for $T \geq 2$, as compared to both the method error and the Monte-Carlo error.

In our case the function $[D(\bar{\rho}(T)/T)]^{1/2}$ tends to zero with rate $1/\sqrt{T}$. So the Monte-Carlo error is proportional to $1/\sqrt{TN}$. Therefore, to reduce the Monte-Carlo error we can increase either N or T . As far as the computational costs are concerned, it does not matter whether we increase N or T . In our case Talay's approach requires the same computational costs as the simulation of Lyapunov exponents by a Monte-Carlo technique. But using Monte-Carlo simulations we not only find $E \bar{\rho}(T)/T$, but also $D(\bar{\rho}(T)/T)$, which is useful for estimating errors.

We simulate the system (9.1) by four different weak schemes: (i) method (7.1) with error $O(h^2 + \varepsilon^2 h)$, which is the simplest method among the schemes proposed in this paper, (ii) method (9.7) with error $O(h^2 + \varepsilon^4 h)$, (iii) the standard method (3.6) with error $O(h^2)$, and (iv) the semi-Runge-Kutta scheme (5.10) with error $O(h^4 + \varepsilon^4 h^2)$, which is the most accurate scheme (in the sense of product $\varepsilon^i h^j$) among the methods we proposed in this paper for a general system with small noise. Method (ii) for system (1.1) has the form

$$(9.7) \quad \begin{aligned} X_{k+1} = & X_k + \varepsilon h^{1/2} \sum_{r=1}^q (\sigma_r \xi_r)_k + h(a + \varepsilon^2 b)_k + \varepsilon h^{3/2} \sum_{r=1}^q (L_1 \sigma_r \xi_r)_k / 2 \\ & + \varepsilon h^{3/2} \sum_{r=1}^q (\Lambda_r a \xi_r)_k / 2 + h^2 (L_1(a + \varepsilon^2 b))_k / 2 + \varepsilon^2 h^2 (L_2 a)_k / 2, \end{aligned}$$

where the random variables ξ_r are simulated using

$$P(\xi = -1) = P(\xi = 1) = \frac{1}{2}.$$

From Table 1 and Figures 1, 2 we can infer that the proposed methods for a system with small noise require less computational effort than the standard ones.

From Figure 1 one can also deduce that the methods $O(h^2 + \dots)$ with time-step $h = \varepsilon^{1/2}$ (i.e., their errors are $O(\varepsilon)$) give similar results. Method (7.1) gives rise to

TABLE 1

Computational results for the Lyapunov exponent $\bar{\lambda}(T)$ for $a = -2$, $c = 1$, $b_1 = b_2 = 2$, $d_1 = 1$, $d_2 = -1$, $\varepsilon = 0.2$, $X^1(0) = 0$, $X^2(0) = 1$, $T = 10$, and for various steps h averaged over N realizations, where $N = 4 \cdot 10^4$ for the methods $O(h^2 + \dots)$ and $N = 1 \cdot 10^6$ for the method $O(h^4 + \varepsilon^4 h^2)$. The exact solution is $\lambda = -2.12$.

	$\frac{1}{N} \sum_{m=1}^q \bar{\rho}^{(m)}(T)/T \pm \frac{2}{\sqrt{N}} \left(\frac{1}{N} \sum_{m=1}^q [\bar{\rho}^{(m)}(T)/T]^2 - \left[\frac{1}{N} \sum_{m=1}^q \bar{\rho}^{(m)}(T)/T \right]^2 \right)^{1/2}$			
h	$O(h^2 + \varepsilon^2 h)$	$O(h^2 + \varepsilon^4 h)$	$O(h^2)$	$O(h^4 + \varepsilon^4 h^2)$
0.3	-2.461 ± 0.004	-2.067 ± 0.002	-2.067 ± 0.002	-2.1228 ± 0.0004
0.2	-2.290 ± 0.003	-2.106 ± 0.002	-2.097 ± 0.002	-2.1195 ± 0.0004
0.1	-2.186 ± 0.002	-2.1198 ± 0.0018	-2.1140 ± 0.0017	-2.1192 ± 0.0004
0.05	-2.150 ± 0.002	-2.1219 ± 0.0018	-2.1186 ± 0.0018	-2.1197 ± 0.0004

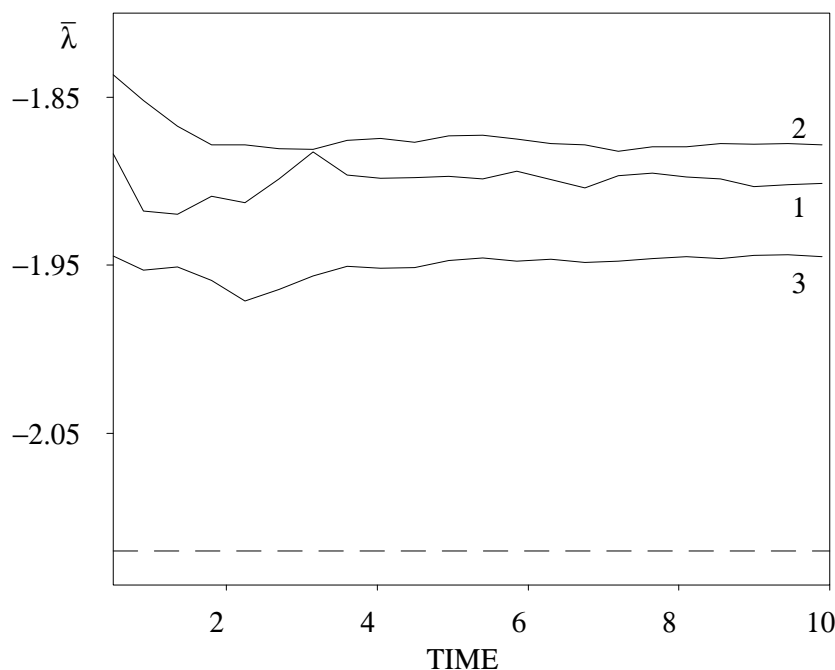


FIG. 1. Time dependence of the function $\bar{\lambda}(T) = E\bar{\rho}(T)/T$ for time-step $h = 0.45$. The other parameters are the same as in Table 1. The solution of the system (9.1), (9.4) is approximated by (1) the method (7.1), (2) the method (9.7), (3) the standard method (3.6). The dashed line shows the exact value of the Lyapunov exponent λ ($\lambda = -2.12$). The number of realizations is $N = 400$, which ensures that the Monte-Carlo errors at $T \geq 7$ are not greater than 0.04 for curve 1 and not greater than 0.02 for curves 2, 3, and they are less than the method errors.

the lowest computational costs, and it is therefore preferable. As can be seen from Figure 2, methods (ii) and (iii) for $h = \varepsilon$ and method (iv) for $h = \varepsilon^{1/2}$ (the errors of these methods are bounded by $O(\varepsilon^2)$) give similar results. Obviously, in this case we prefer semi-Runge-Kutta method (iv) because it saves CPU time.

The data of Table 1 show that the methods $O(h^2 + \varepsilon^2 h)$, $O(h^2 + \varepsilon^4 h)$, and $O(h^2)$ can be improved by using the Talay-Tubaro expansion (see section 7). For $\varepsilon = 0.2$, for instance, one can calculate $C_1(\varepsilon)$ and $C_2(\varepsilon)$ from the expansion of the global error

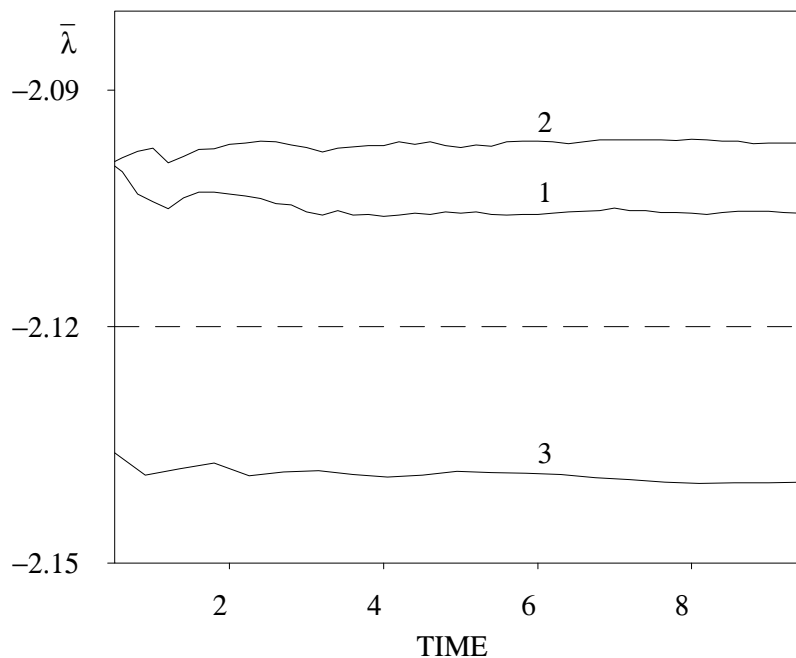


FIG. 2. Time dependence of the function $\bar{\lambda}(T) = E\bar{\rho}(T)/T$ simulated using (1) method (9.7) with $h = 0.2$, (2) the standard method (3.6) with $h = 0.2$, (3) method (5.10) with $h = 0.45$. The other parameters are the same as in Table 1. The dashed line shows the exact value of the Lyapunov exponent λ ($\lambda = -2.12$). The Monte-Carlo errors at $T \geq 7$ are not greater than 0.002 ($N = 40000$) and are less than the method errors.

of method (7.1) (see section 7, Theorem 7.1) to obtain $C_1(\varepsilon) \approx 2.1$ and $C_2(\varepsilon) \approx 10.2$. Let us emphasize that if some constants in the error expansion have opposite signs, then the error will become a nonmonotonous function of the time-step h and thus may increase while h is decreasing. Such behavior is demonstrated in Table 1 (see the methods $O(h^2 + \varepsilon^4 h)$ and $O(h^4 + \varepsilon^4 h^2)$).

In Figure 3 we show the time dependence of the function $\bar{\rho}(T)/T$ when taking Talay's approach to calculate Lyapunov exponents, i.e., along a single trajectory of a weak scheme. One can see that in this case our methods give accurate results and allow us to reduce computational costs. By the Monte-Carlo simulations using method (5.10) with $h = 0.2$, $T = 10$, $N = 10^6$ we achieve an accuracy of $\approx 0.5 \cdot 10^{-3}$ (see Table 1).

Throughout our tests we use procedure RAN1 from [13] to generate uniform random numbers.

Remark. Note that the function $\ln|x|$ does not belong to the class **F**. Thus, if $\lambda > 0$, to deduce strict conclusions one can consider the function $\ln(1 + |x|)$ instead of $\ln|x|$. The function $\ln(1 + |x|)$ already belongs to the class **F** and $\lim_{t \rightarrow \infty} \ln(1 + |X(t)|)/t = \lim_{t \rightarrow \infty} \ln(|X(t)|)/t$. As can be seen by carrying out numerical tests, simulations of the function $\ln(1 + |x|)$ yield the same results as simulations of $\ln|x|$. For $\lambda < 0$ one can either switch to the function $\ln(1 + 1/|x|)$ or to the system

$$(9.8) \quad dX = (\gamma I + AX)dt + \varepsilon \sum_{r=1}^q B_r X dW_r$$

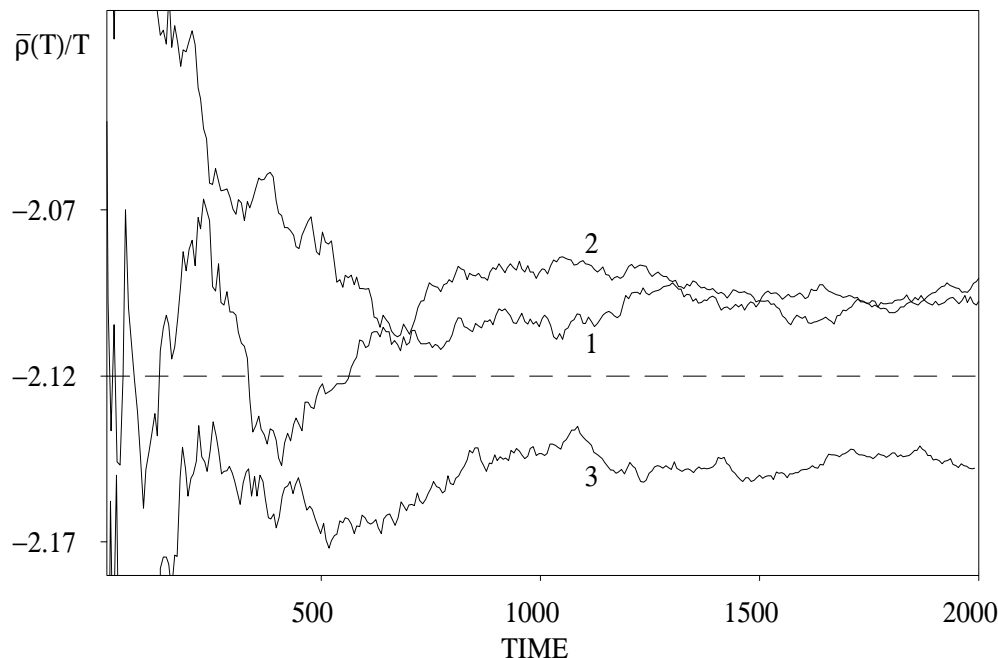


FIG. 3. Time dependence of the function $\bar{\rho}(T)/T$ computed along a single trajectory using (1) method (9.7) with $h = 0.1$, (2) the standard method (3.6) with $h = 0.1$, (3) method (5.10) with $h = 0.3$. The other parameters are the same as in Table 1. The dashed line shows the exact value of the Lyapunov exponent λ ($\lambda = -2.12$).

instead of system (9.1). The Lyapunov exponent of system (9.8) is equal to $\gamma + \lambda$, and if we choose γ , such that $\gamma + \lambda > 0$, we can use the function $\ln(1 + |x|)$ again.

10. Appendix. Derivation of the error estimate for the Runge–Kutta method (5.9). We introduce the notation

$$\Delta = X(t+h) - x, \quad \bar{\Delta} = \bar{X} - x, \quad \tilde{\Delta} = \tilde{X} - x, \quad X(t) = \bar{X}(t) = \tilde{X} = x,$$

where \bar{X} is the one-step approximation corresponding to the semi-Runge–Kutta method (5.9) and \tilde{X} is the one-step approximation corresponding to the Taylor-type method (4.5).

We shall prove the inequalities

$$(10.1) \quad \left| E \left(\prod_{j=1}^m \bar{\Delta}^{i_j} - \prod_{j=1}^m \tilde{\Delta}^{i_j} \right) \right| \leq K(h^5 + \varepsilon^2 h^3), \quad i_j = 1, \dots, n, \quad m = 1, \dots, 5,$$

$$(10.2) \quad E \prod_{j=1}^6 |\bar{\Delta}^{i_j}| \leq K(h^5 + \varepsilon^2 h^3), \quad i_j = 1, \dots, n.$$

From these inequalities the error of the method (5.9) can easily be read off.

The one-step approximation \tilde{X} can be estimated in the same manner as was done for the one-step error of approximation (3.12) in section 3. For the approximation \tilde{X}

we can derive

$$(10.3) \quad \left| E \left(\prod_{j=1}^m \Delta^{i_j} - \prod_{j=1}^m \tilde{\Delta}^{i_j} \right) \right| \leq K(h^5 + \varepsilon^2 h^3), \quad i_j = 1, \dots, n, \quad m = 1, \dots, 5.$$

Then, as in section 3, from (10.1)–(10.3) we obtain an estimate of the local error of the approximation \bar{X} in the form

$$(10.4) \quad |Ef(X(t+h)) - Ef(\bar{X}(t+h))| \leq K(h^5 + \varepsilon^2 h^3).$$

From (10.4) and Theorem 2.1 it follows that the global error of the semi-Runge–Kutta method (5.9) can be estimated by

$$(10.5) \quad |Ef(X(t_k)) - Ef(X_k)| \leq K(h^4 + \varepsilon^2 h^2)$$

for all N and $k = 0, 1, \dots, N$.

Now we prove the inequalities (10.1) and (10.2). The inequalities (10.2) evidently follow from the form of the one-step approximation \bar{X} by an argument similar to the proof of inequalities (3.2) in section 3.

To prove (10.1) we expand the terms of \bar{X} in powers of h at the point (t, x) . The first term is

$$(10.6) \quad \varepsilon h^{1/2}(\sigma_r(t, x) + \sigma_r(t+h, x))\xi_r/2 = \varepsilon h^{1/2}\sigma_r + \varepsilon h^{3/2}\frac{\partial\sigma_r}{\partial t}\xi_r/2 + \rho_{1r},$$

where

$$\rho_{1r} = \varepsilon h^{5/2}\frac{\partial^2\sigma_r}{\partial t^2}\xi_r/4 + O(\varepsilon h^{7/2})\xi_r.$$

It is clear that ρ_{1r} has the properties

$$E\rho_{1r} = 0, \quad r = 1, \dots, q, \quad |E(\rho_{1r}\rho_{1s})| \leq K\varepsilon^2 h^5, \quad r, s = 1, \dots, q,$$

$$(10.7) \quad |E(\rho_{1r} \cdot \varepsilon h^{1/2}\xi_s)| \leq K\varepsilon^2 h^3, \quad r, s = 1, \dots, q.$$

The term

$$B \equiv \frac{\varepsilon^2 h}{2} \left\{ b(t, x) + b \left(t + h, x + \varepsilon h^{1/2} \sum_{j=1}^q \sigma_j \xi_j + h(a + \varepsilon^2 b) \right) \right\}$$

has the expansion

$$(10.8) \quad B = \frac{\varepsilon^2 h}{2} \left\{ 2b + h \frac{\partial b}{\partial t} + \varepsilon h^{1/2} \sum_{j=1}^q \sum_{i=1}^n \sigma_j^i \xi_j \frac{\partial b}{\partial x^i} + h \sum_{i=1}^n (a + \varepsilon^2 b)^i \frac{\partial b}{\partial x^i} \right. \\ \left. + \frac{\varepsilon^2 h}{2} \sum_{j=1}^q \sum_{s=1}^q \sum_{i=1}^n \sum_{l=1}^n \sigma_j^i \sigma_s^l \xi_j \xi_s \frac{\partial^2 b}{\partial x^i \partial x^l} \right\} + A + O(\varepsilon^2 h^3),$$

where A contains terms of the form $A_{jsl}\varepsilon^5 h^{5/2}\xi_i \xi_j \xi_l$, $A_j \varepsilon^3 h^{5/2}\xi_j$ and A_{jsl} and A_j are constants containing a , b , σ_r , and their derivatives at the point (t, x) .

In the braces of (10.8) we replace the term

$$\frac{1}{2} \sum_{j=1}^q \sum_{s=1}^q \sum_{i=1}^n \sum_{l=1}^n \sigma_j^i \sigma_s^l \xi_j \xi_s \frac{\partial^2 b}{\partial x^i \partial x^l}$$

by

$$\frac{1}{2} E \sum_{j=1}^q \sum_{s=1}^q \sum_{i=1}^n \sum_{l=1}^n \sigma_j^i \sigma_s^l \xi_j \xi_s \frac{\partial^2 b}{\partial x^i \partial x^l} = \frac{1}{2} \sum_{j=1}^q \sum_{i=1}^n \sum_{l=1}^n \sigma_j^i \sigma_j^l \frac{\partial^2 b}{\partial x^i \partial x^l}.$$

Then formula (10.8) becomes

$$(10.9) \quad B = \varepsilon^2 h \left\{ 2b + h \frac{\partial b}{\partial t} + \varepsilon h^{1/2} \sum_{j=1}^q \sum_{i=1}^n \sigma_j^i \xi_j \frac{\partial b}{\partial x^i} + h \sum_{i=1}^n a^i \frac{\partial b}{\partial x^i} + \varepsilon^2 h L_2 b \right\} / 2 + \rho_2,$$

where

$$\rho_2 = \frac{\varepsilon^2 h}{2} \left\{ -\frac{\varepsilon^2 h}{2} \sum_{j=1}^q \sum_{i=1}^n \sum_{l=1}^n \sigma_j^i \sigma_j^l \frac{\partial^2 b}{\partial x^i \partial x^l} + \frac{\varepsilon^2 h}{2} \sum_{j=1}^q \sum_{s=1}^q \sum_{i=1}^n \sum_{l=1}^n \sigma_j^i \sigma_s^l \xi_j \xi_s \frac{\partial^2 b}{\partial x^i \partial x^l} \right\} + A + O(\varepsilon^2 h^3).$$

For the remainder ρ_2 , we have

$$(10.10) \quad \begin{aligned} |E\rho_2| &\leq K\varepsilon^2 h^3, \quad |E\rho_2^2| \leq K(\varepsilon^4 h^6 + \varepsilon^8 h^4), \quad |E(\rho_2 \cdot \varepsilon h^{1/2} \xi_m)| \leq K\varepsilon^4 h^3, \\ |E(\rho_2 \cdot \varepsilon^2 h \xi_{mr})| &\leq K(\varepsilon^4 h^4 + \varepsilon^6 h^3), \quad |E(\rho_2 \cdot \varepsilon h^{1/2} \xi_m \cdot \varepsilon h^{1/2} \xi_r)| \leq K(\varepsilon^4 h^4 + \varepsilon^6 h^3), \\ |E(\rho_{1,r} \rho_2)| &\leq K\varepsilon^4 h^5, \quad m, r = 1, \dots, q. \end{aligned}$$

For k_i , $i = 1, \dots, 4$, of (5.6) we have

$$(10.11) \quad \begin{aligned} \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) &= ha + \frac{1}{2}h^2 L_1 a + \frac{1}{6}h^3 L_1^2 a + \frac{1}{24}h^4 L_1^3 a \\ &+ \frac{1}{2}\varepsilon h^{3/2} \sum_{r=1}^q \sum_{i=1}^n \sigma_r^i \xi_r \frac{\partial a}{\partial x^i} + \frac{1}{2}\varepsilon^2 h^2 L_2 a + \rho_3, \end{aligned}$$

where

$$\rho_3 = -\frac{\varepsilon^2 h^2}{4} \sum_{j=1}^q \sum_{i=1}^n \sum_{l=1}^n \sigma_j^i \sigma_j^l \frac{\partial^2 a}{\partial x^i \partial x^l}$$

$$(10.12) \quad + \frac{\varepsilon^2 h^2}{4} \sum_{r=1}^q \sum_{s=1}^q \sum_{i=1}^n \sum_{j=1}^n \sigma_r^i \sigma_s^j \xi_r \xi_s \frac{\partial^2 a}{\partial x^i \partial x^j} + A + O(h^5 + \varepsilon^2 h^3).$$

The expression A of (10.12) contains terms of the form $\varepsilon^3 h^{5/2} A_{rsj} \xi_r \xi_s \xi_j$, $\varepsilon h^{5/2} A_r \xi_r$, where A_{rsj} and A_r are constants containing a , b , σ_r , and their derivatives at the point (t, x) .

The remainder ρ_3 has the properties

$$\begin{aligned} |E\rho_3| &\leq K(h^5 + \varepsilon^2 h^3), & |E\rho_3^2| &\leq K(h^{10} + \varepsilon^2 h^5 + \varepsilon^4 h^4), \\ |E(\rho_3 \cdot \varepsilon h^{1/2} \xi_m)| &\leq K\varepsilon^2 h^3, & |E(\rho_3 \cdot \varepsilon^2 h \xi_{mr})| &\leq K\varepsilon^4 h^3, \\ |E(\rho_3 \cdot \varepsilon h^{1/2} \xi_m \cdot \varepsilon h^{1/2} \xi_r)| &\leq K\varepsilon^4 h^3, & |E(\rho_3 \rho_{1r})| &\leq K\varepsilon^2 h^5, \\ (10.13) \quad |E(\rho_3 \rho_2)| &\leq K(\varepsilon^4 h^5 + \varepsilon^6 h^4), & m, r &= 1, \dots, q. \end{aligned}$$

From (10.6), (10.9), (10.11) and from the forms of the one-step approximations \bar{X} and \tilde{X} , we get

$$(10.14) \quad \bar{\Delta} - \tilde{\Delta} = \sum_{r=1}^q \rho_{1r} + \rho_2 + \rho_3.$$

Then from (10.7), (10.10), (10.13), and (10.14) we obtain (10.1).

11. Appendix. Proof of Theorem 7.1. Consider (cf. (8.6)) the following:

$$(11.1) \quad u(s, x) = Ef(X_{s,x}^\varepsilon(T)) = u^\varepsilon(s, x) = u^o(s, x) + \varepsilon^2 u^1(s, x; \varepsilon).$$

As in the proof of the weak convergence theorem (see [7, 8]), we obtain the equality

$$\begin{aligned} R &\equiv Ef(X^\varepsilon(T)) - Ef(\bar{X}^\varepsilon(T)) = Ef(X^\varepsilon(t_N)) - Ef(\bar{X}_N^\varepsilon) \\ &= \sum_{i=0}^{N-1} \left\{ Eu(t_{i+1}, X_{t_i, \bar{X}_i}^\varepsilon(t_{i+1})) - Eu(t_{i+1}, \bar{X}_{t_i, \bar{X}_i}^\varepsilon(t_{i+1})) \right\} \\ (11.2) \quad &= E \sum_{i=0}^{N-1} E(u(t_{i+1}, X_{t_i, \bar{X}_i}^\varepsilon(t_{i+1})) - u(t_{i+1}, \bar{X}_{t_i, \bar{X}_i}^\varepsilon(t_{i+1})) | \bar{X}_i^\varepsilon). \end{aligned}$$

From [8] we get

$$Eu(t+h, X_{t,x}(t+h)) = u + hLu + \frac{1}{2}h^2L^2u + \frac{1}{6}h^3L^3u + O(h^4),$$

and taking into account that $L = L_1 + \varepsilon^2 L_2$, we write

$$\begin{aligned} Eu(t+h, X_{t,x}(t+h)) &= u + h(L_1 + \varepsilon^2 L_2)u \\ (11.3) \quad &+ \frac{1}{2}h^2(L_1 + \varepsilon^2 L_2)^2u + \frac{1}{6}h^3L_1^3u + O(h^4 + \varepsilon^2 h^3). \end{aligned}$$

At a single step the method (7.1) has the form

$$(11.4) \quad \bar{X} = x + \varepsilon h^{1/2} \sum_{r=1}^q \sigma_r \xi_r + h(a + \varepsilon^2 b) + \frac{1}{2} h^2 L_1 a,$$

$$P(\xi = -1) = P(\xi = 1) = \frac{1}{2}.$$

Expanding $Eu(t + h, \bar{X}_{t,x}(t + h))$ in powers of h (by using the Taylor formula), we obtain

$$(11.5) \quad \begin{aligned} Eu(t + h, \bar{X}_{t,x}(t + h)) &= u + h(L_1 + \varepsilon^2 L_2)u \\ &+ \frac{1}{2} h^2 L_1^2 u + \varepsilon^2 h^2 A_2 + \varepsilon^4 h^2 A_3 + h^3 A_1 + O(h^4 + \varepsilon^2 h^3), \end{aligned}$$

where

$$(11.6) \quad \begin{aligned} A_1 &= \frac{1}{2} \sum_{i=1}^n (L_1 a)^i \frac{\partial^2 u}{\partial t \partial x^i} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a^i (L_1 a)^j \frac{\partial^2 u}{\partial x^i \partial x^j} + \frac{1}{6} \frac{\partial^3 u}{\partial t^3} \\ &\quad + \frac{1}{6} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n a^i a^j a^l \frac{\partial^3 u}{\partial x^i \partial x^j \partial x^l}, \\ A_2 &= \sum_{i=1}^n b^i \frac{\partial^2 u}{\partial t \partial x^i} + \sum_{i=1}^n \sum_{j=1}^n b^i a^j \frac{\partial^2 u}{\partial x^i \partial x^j} + \frac{1}{2} \sum_{r=1}^q \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sigma_r^i \sigma_r^j a^l \frac{\partial^3 u}{\partial x^i \partial x^j \partial x^l} \\ &\quad + \frac{1}{2} \sum_{r=1}^q \sum_{i=1}^n \sum_{j=1}^n \sigma_r^i \sigma_r^j \frac{\partial^3 u}{\partial t \partial x^i \partial x^j}, \\ A_3 &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n b^i b^j \frac{\partial^2 u}{\partial x^i \partial x^j} + \frac{1}{2} \sum_{r=1}^q \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sigma_r^i \sigma_r^j b^l \frac{\partial^3 u}{\partial x^i \partial x^j \partial x^l} \\ &\quad + \frac{1}{24} \sum_{r=1}^q \sum_{s=1}^q \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{m=1}^n \sigma_r^i \sigma_r^j \sigma_s^l \sigma_s^m \frac{\partial^4 u}{\partial x^i \partial x^j \partial x^l \partial x^m}. \end{aligned}$$

All coefficients and derivatives in (11.5) and (11.6) are calculated at (t, x) and depend on ε . By (11.3) and (11.5) we have

$$(11.7) \quad \begin{aligned} &Eu(t + h, X_{t,x}(t + h)) - Eu(t + h, \bar{X}_{t,x}(t + h)) \\ &= \varepsilon^2 h^2 B_2 + \varepsilon^4 h^2 B_3 + h^3 B_1 + O(h^4 + \varepsilon^2 h^3), \end{aligned}$$

where

$$(11.8) \quad B_1 = \frac{1}{6} L_1^3 u - A_1, \quad B_2 = \frac{1}{2} L_1 L_2 u + \frac{1}{2} L_2 L_1 u - A_2, \quad B_3 = \frac{1}{2} L_2^2 u - A_3.$$

Substituting the conditional variant of (11.7) in (11.2) we obtain

$$(11.9) \quad R = E \sum_{i=0}^{N-1} \{B_2(\varepsilon; t_i, \bar{X}_i^\varepsilon) \varepsilon^2 h^2 + B_3(\varepsilon; t_i, \bar{X}_i^\varepsilon) \varepsilon^4 h^2 + B_1(\varepsilon; t_i, \bar{X}_i^\varepsilon) h^3\} + O(h^3 + \varepsilon^2 h^2).$$

For each $j = 1, 2, 3$ consider the $(n+1)$ -dimensional system

$$(11.10) \quad dX = (a + \varepsilon^2 b) dt + \sum_{r=1}^q \sigma_r dW_r, \quad X(t_o) = X_o, \\ dY = B_j dt, \quad Y(t_o) = 0,$$

where the first n equations are the original ones (see (1.1)) and the last equation describes Y .

Solving the system (11.10) by method (7.1) we obtain

$$E \int_{t_o}^T B_j(\varepsilon; \vartheta, X^\varepsilon(\vartheta)) d\vartheta = EY(T) \\ = E \left\{ \sum_{i=0}^{N-1} \left[B_j(\varepsilon; t_i, \bar{X}_i^\varepsilon) h + \frac{1}{2} L_1 B_j(\varepsilon; t_i, \bar{X}_i^\varepsilon) h^2 \right] \right\} + O(h^2 + \varepsilon^2 h) \\ = E \sum_{i=0}^{N-1} B_j(\varepsilon; t_i, \bar{X}_i^\varepsilon) h + O(h).$$

Therefore,

$$(11.11) \quad E \sum_{i=0}^{N-1} B_j(\varepsilon; t_i, \bar{X}_i^\varepsilon) h = E \int_{t_o}^T B_j(\varepsilon; \vartheta, X^\varepsilon(\vartheta)) d\vartheta + O(h).$$

Substituting (11.11) in (11.9) we obtain

$$(11.12) \quad R = C_1(\varepsilon) h^2 + C_2(\varepsilon) \varepsilon^2 h + O(h^3 + \varepsilon^2 h^2),$$

where

$$(11.13) \quad C_1(\varepsilon) = E \int_{t_o}^T B_1(\varepsilon; \vartheta, X^\varepsilon(\vartheta)) d\vartheta, \\ C_2(\varepsilon) = E \int_{t_o}^T B_2(\varepsilon; \vartheta, X^\varepsilon(\vartheta)) d\vartheta + \varepsilon^2 E \int_{t_o}^T B_3(\varepsilon; \vartheta, X^\varepsilon(\vartheta)) d\vartheta.$$

The coefficients $C_i(\varepsilon)$ do not depend on the time increment h , but depend on ε (see (11.1)) as

$$(11.14) \quad C_i(\varepsilon) = C_i^o + O(\varepsilon^2).$$

The proof of Theorem 7.1 is complete.

Acknowledgment. We wish to express our gratitude to the Deutscher Akademischer Austauschdienst for supporting M. V. Tret'yakov with a Forschungsstipendium aus Sondermitteln der NATO.

REFERENCES

- [1] L. ARNOLD AND V. WIHSTUTZ, EDS., *Lyapunov Exponents*, Lecture Notes in Mathematics 1186, Springer-Verlag, Berlin, 1986, pp. 1–26.
- [2] E. I. AUSLENDER AND G. N. MILSTEIN, *Asymptotic expansions of the Lyapunov index for linear stochastic systems with small noises*, Prikl. Mat. Mekh., 46 (1982), pp. 358–365.
- [3] M. I. FREIDLIN AND A. D. WENTZELL, *Random Perturbations of Dynamical Systems*, Springer-Verlag, Berlin, 1984.
- [4] R. Z. HASMINSKI, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Gröninger, the Netherlands, 1980.
- [5] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1992.
- [6] G. N. MILSTEIN, *A method of second-order accuracy integration of stochastic differential equations*, Theory Probab. Appl., 23 (1978), pp. 396–401.
- [7] G. N. MILSTEIN, *Weak approximation of solutions of systems of stochastic differential equations*, Theory Probab. Appl., 30 (1985), pp. 750–766.
- [8] G. N. MILSTEIN, *Numerical Integration of Stochastic Differential Equations*, Ural University Press, Sverdlovsk, Russia, 1988; Kluwer Academic Publishers, Norwell, MA, 1995.
- [9] G. N. MILSTEIN AND M. V. TRET'YAKOV, *Mean-square numerical methods for stochastic differential equations with small noises*, SIAM J. Sci. Comput., 18 (1997), pp. 1067–1087.
- [10] G. N. MILSTEIN AND M. V. TRET'YAKOV, *Numerical solution of differential equations with colored noise*, J. Stat. Phys., 77 (1994), pp. 691–715.
- [11] G. N. MILSTEIN AND M. V. TRET'YAKOV, *Weak approximation for stochastic differential equations with small noises*, Preprint No. 123, Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin, 1994.
- [12] E. PARDOUX AND D. TALAY, *Discretization and simulation of stochastic differential equations*, Acta Appl. Math., 3 (1985), pp. 23–47.
- [13] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY, *Numerical Recipes in Fortran: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 1992.
- [14] D. TALAY, *Efficient numerical schemes for the approximation of expectations of functionals of the solution of an SDE and applications*, Lecture Notes in Control and Inform. Sci. 61, Springer-Verlag, Berlin, 1984, pp. 294–313.
- [15] D. TALAY, *Approximation of upper Lyapunov exponents of bilinear stochastic differential equations*, SIAM J. Numer. Anal., 28 (1991), pp. 1141–1164.
- [16] D. TALAY AND L. TUBARO, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Anal. Appl., 8 (1990), pp. 483–509.
- [17] W. WAGNER AND E. PLATEN, *Approximation of Ito Integral Equations*, preprint, Zentralinstitut für Mathematik und Mechanik, Akad. der Wiss. der DDR, Berlin, 1978.